DTU Library

# Statistical Modelling of Wind Proles - Data Analysis and Modelling

**Jónsson, Tryggvi; Pinson, Pierre**

*Publication date:*
2013

*Document Version*
Publisher's PDF, also known as Version of record

# Statistical Modelling of Wind Profiles - Data Analysis and Modelling

Tryggvi Jónsson & Pierre Pinson

January 17, 2013

## 1 Introduction

The aim of the analysis presented in this document is to investigate whether statistical models can be used to make very short-term predictions of wind profiles. The data set used for this analysis is the one described in Mann et al. (2010).

The available data set consists of 1 Hz measurements of wind speeds and direction, measured by the Østerild LIDAR, during the period from February 1st 2010 until May 5th 2011. Due to the huge amount of data logged every second, most of the results presented in this report are only based on the first two months of the data set.

The measurements are taken by three different LIDARs and at 10 different heights. The LIDARs are placed at the Østerild test site for wind turbines, located in the Northwestern part of Denmark. Every second, each of the LIDAR cubes logs the wind speed and direction at 10 different altitudes, between 45 meters and 300 meters above sea level. The measurement heights are:

$$\begin{bmatrix} 45 & 60 & 80 & 100 & 120 & 140 & 170 & 200 & 250 & 300 \end{bmatrix} meters.$$

Based on these data a wind profile using the *wind profile power law* on the form:

$$u_x = u_r \left(\frac{x}{r}\right)^{\alpha}, \tag{1}$$

is estimated, where $u_x$ is the wind speed (in m/s) at height $x$, $u_r$ is the (known) wind speed at a reference height $r$[1]. The $\alpha$'s are sampled with a time resolution of 30 *secs*, or 0.033 Hz, under a least squares criteria using the 45 meter measurement is used as a reference height. Further introduction to the data set is given in the next section.

Based on this data set, statistical models for a single step forecasting the the $\alpha$-parameter are constructed and their ability to predict the wind profile is assessed. All analysis and forecasting is focused on the $\alpha$-parameter for the North LIDAR cube, while the others (the West and the South ones) are used as explanatory variables in some models. It is quite plausible though that similar results as those listed here can be obtained for the other two cubes.

The remainder of this report is structured as follows: a more comprehensive description of the data and data analysis are presented in Section 2, followed by a theoretical outline of the models used in Section 3. The various model setups tried and empirical results are the subject of Section 4 while conclusions and concluding remarks are given in Section 5. Finally, some additional analysis and results are shown in Appendices A and B.

[1] The heights are sometimes noted $z_x$ and $z_r$ but it seems weird since $z_r = r$ and $z_x = x$

## 2 The Data and its Quality

During the estimation process, the following variables are logged besides the estimated $\alpha$-parameter:

1. The average wind speed during the 30 second period at each height, $\overline{U}_t^{(h)}$.
2. The average wind direction during the 30 second period at each height, $\overline{D}_t^{(h)}$.
3. The number of available observations during the 30 second period at each height, $N_{Av}^{(h)}(t)$.

Overall, the data availability is quite good, meaning that there are relatively few observations missing. Most of the data that is missing entirely, during this two month period, are consecutive missing observations during February 20th - February 23rd. However, the number of missing observation is significantly larger for the two highest measuring altitudes, 250 and 300 meters. As it turns out, missing the high altitude measurements seems to result in unstable estimates of the $\alpha$-parameter which appears in the corresponding time series as heteroskedacity. Likewise low wind speed seems to trigger the same effect. Both of these properties of the time series will be illustrated further in the following.

In Figure 1, the time-series $\alpha$ is plotted with 5 different color labeling. Linecolours in the first two plots in the figure are assigned according to the data availability at 300 m during time period $t$, $N_{Av}^{(300)}(t)$, and the average wind speed, $\overline{U}_t$, across all heights during the same period respectively. In the next two plots, the same variables have been categorized into 2 and 3 categories as follows:

$$\text{Data Availability} = \begin{cases} \text{High} & \text{if} \frac{N_{Av}}{N} \geq 0.5 \\ \text{Low} & \text{if } \frac{N_{Av}}{N} < 0.5 \end{cases}, \quad \text{Wind Speed} = \begin{cases} \text{High} & \text{if } \overline{U} > 10 \\ \text{Medium} & \text{if } 5 \leq \overline{U} \leq 10 \\ \text{Low} & \text{if } \overline{U} < 5 \end{cases}. \tag{2}$$

In the 5th and final plot, the colors are assigned according to both categories simultaneously.

By viewing all 5 plots together, one can see that abnormal $\alpha$-values seem to be limited to times when either the availability of the high altitude observations or the observed wind speed is low or even both. An example of this is shown in Figure 2, which depicts the $\alpha$-series from 18:00 on February 3rd until midnight on February 5th. During this 30 hour period, the variance of $\alpha$ is inflated twice, once around 03:00 on February 4th and again in the evening on that same day. In the first case, low wind speed seems to be the trigger while, in the second case, data availability has decreased.

Taking a time-invariant perspective, the density of the $\alpha$'s is shown in Figure 3, using the categorization of the last plot of Figure 1. Both marginal and stacked variation is shown (thus including the unconditional density) as well as a conditional one. Here, the higher variance of $\alpha$ during times of low wind speed and/or low data availability appears as larger tails of the corresponding conditional distributions.

Due to these observations, the models tried out for the forecasting where fitted in parallel to the full data set and to a subset where data availability at 300 meters was above 50% during the 30 second interval and wind speed was more than 5 $m/s$. The idea behind doing so is to investigate whether the methods can be used for other the more common scenario and subsequently introduce regime switching to the model so that low wind speeds can be accommodated as well.

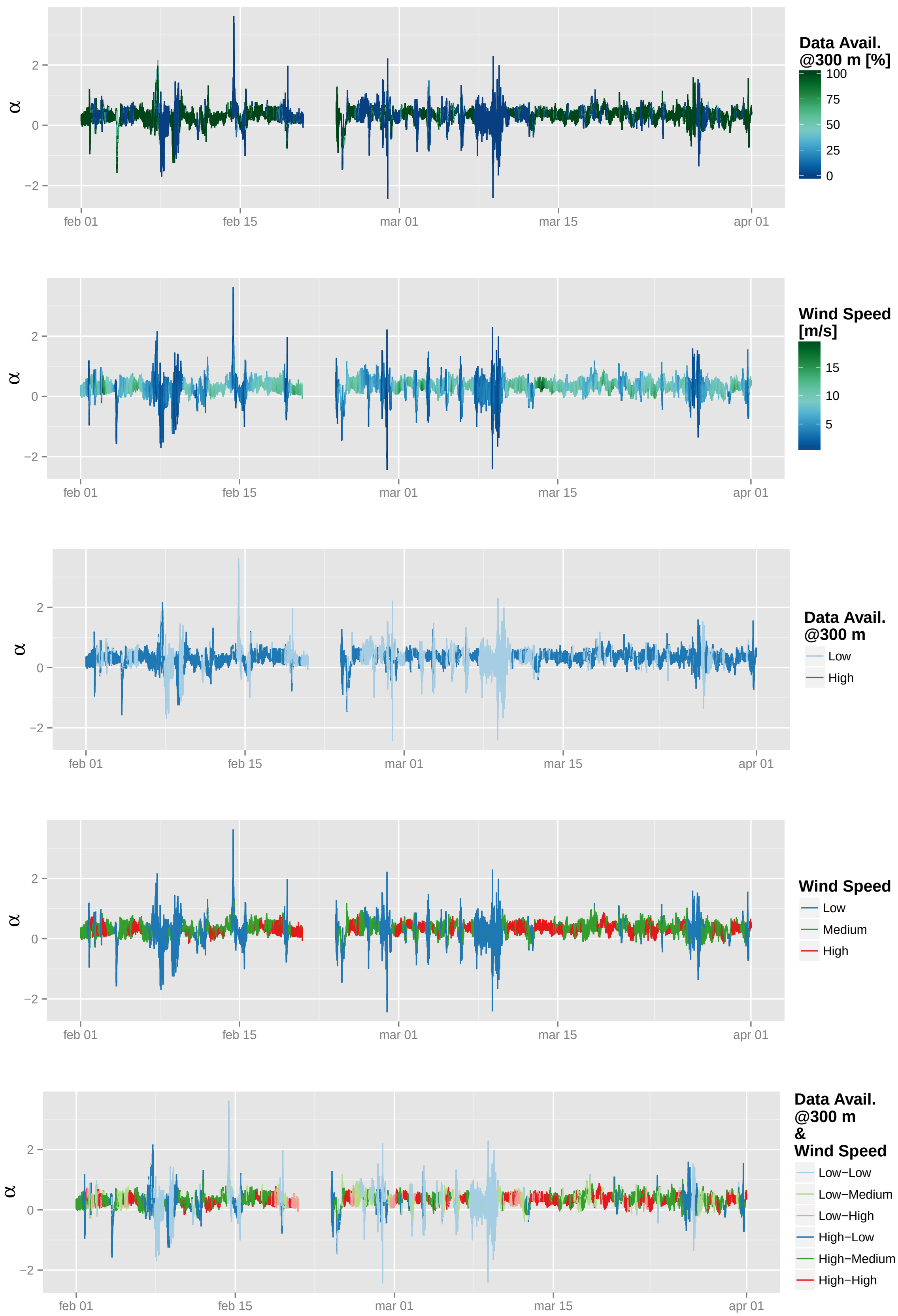
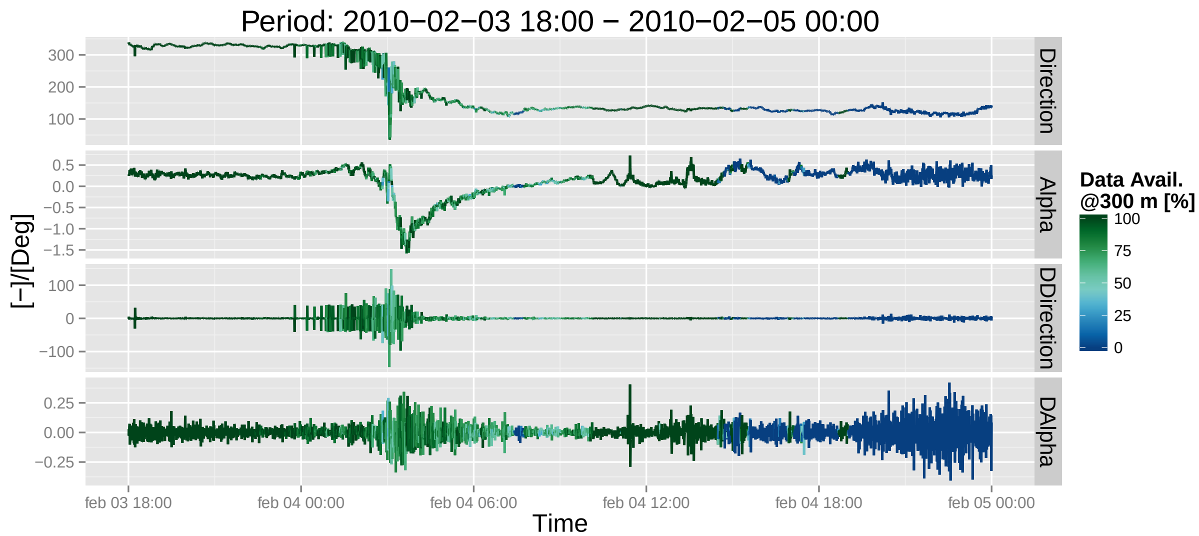
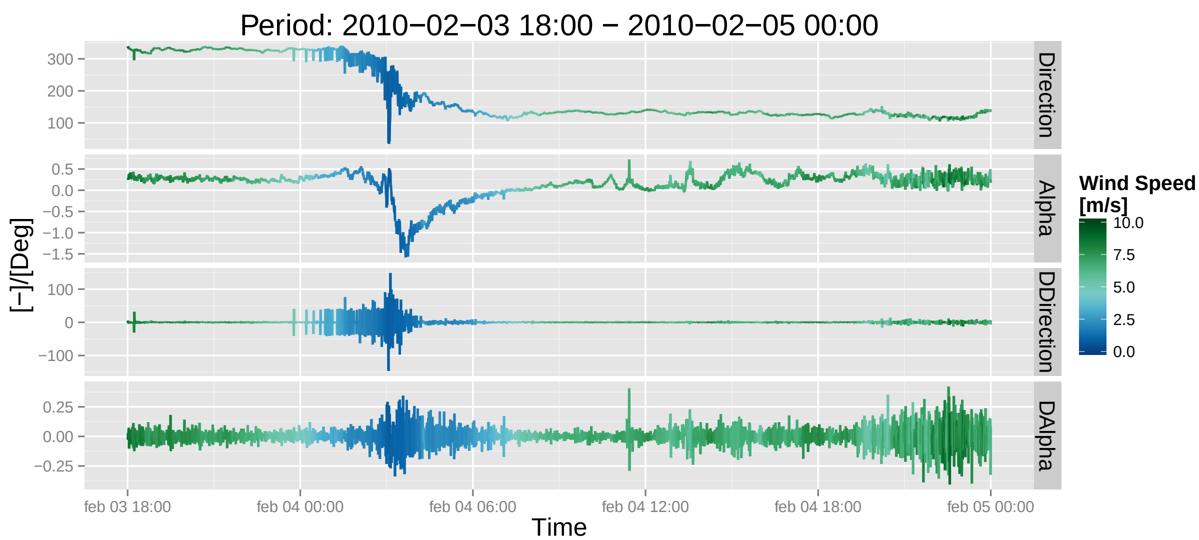
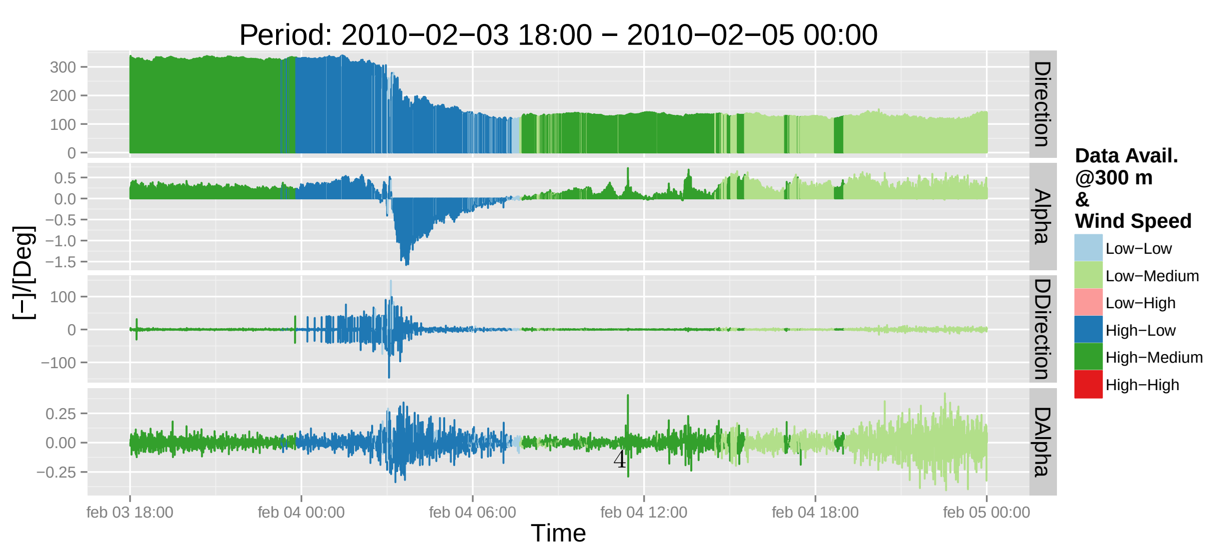





Figure 1: Time series plots of $\alpha$ with different color codes


Period: 2010−02−03 18:00 − 2010−02−05 00:00
Direction
Alpha
DDirection
DAlpha
[−]/[Deg]
Time
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Data Avail.
@300 m [%]
100
75
50
25
0


Period: 2010−02−03 18:00 − 2010−02−05 00:00
Direction
Alpha
DDirection
DAlpha
[−]/[Deg]
Time
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Wind Speed
[m/s]
10.0
7.5
5.0
2.5
0.0


Period: 2010−02−03 18:00 − 2010−02−05 00:00
Direction
Alpha
DDirection
DAlpha
[−]/[Deg]
Time
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Data Avail.
@300 m
&
Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High

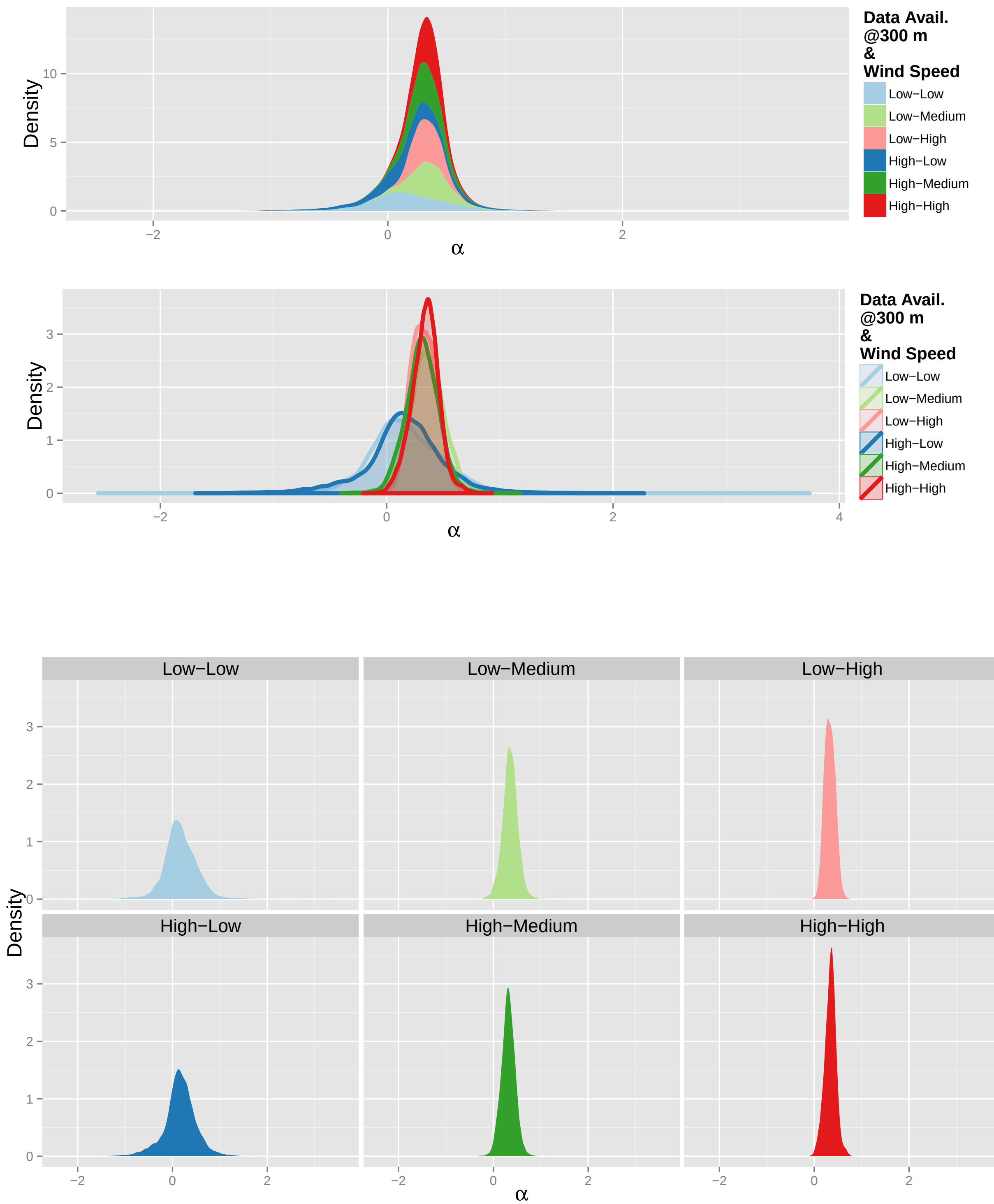



Figure 3: Density plots of the $\alpha$'s, split according to data availability and wind speed categories

# 3 Model Outline

The methodological focus of this report is on the ARIMAX model using recursive parameter estimation. Conditional parametric version of that model is estimated as well as an unconditional one. The main characteristics of such model are briefly outlined in the following. The description provided here is quite superficial, so readers interested in in-depth description of the methods are referred to the references given.

## 3.1 Recursive Least Squares Parameter Estimation

Let $\boldsymbol{Y}_t$ be a stochastic time series, comprising $N$ observations $y_t$, which can be described by the the model

$$y_t + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} = \omega_1 u_{t-1} + \ldots + \omega_s u_{t-s} + \varepsilon_t(\phi,\omega,\theta) + \theta_1 \varepsilon_{t-1}(\phi,\omega,\theta) + \ldots + \theta_q \varepsilon_{t-q}(\phi,\omega,\theta) \quad (3)$$

which is an ARIMAX model of order $(p, d = 0, q, s)$ model. Now define:

$$\boldsymbol{X}_t^T = [y_{t-1}, \ldots, y_{t-p}, u_{t-1}, \ldots, u_{t-s}, \varepsilon_{t-1}(\phi,\omega,\theta), \ldots, \varepsilon_{t-q}(\phi,\omega,\theta))] \quad (4)$$

$$\boldsymbol{\theta}^T = [\phi_1, \ldots, \phi_p y_{t-p}, \omega_1, \ldots, \omega_s, \theta_1, \ldots, \theta_q] \quad (5)$$

which allows us to write the model in (3) as

$$y_t = \boldsymbol{X}_t^T(\boldsymbol{\theta})\boldsymbol{\theta} + \varepsilon_t. \quad (6)$$

An off-line (time invariant) least squares estimate of the parameters, $\widehat{\boldsymbol{\theta}}$ in the model (6) is then found as

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} S_N(\boldsymbol{\theta}) \quad (7)$$

where

$$S_N(\boldsymbol{\theta}) = \sum_{t=1}^{N} \varepsilon_t^2(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{N} \left(y_t - \boldsymbol{X}_t^T(\boldsymbol{\theta})\boldsymbol{\theta}\right)^2. \quad (8)$$

Now allowing the parameter vector $\boldsymbol{\theta}$ to vary over time, so it becomes $\boldsymbol{\theta}_t$, and introducing the forgetting factor $\lambda$ ($\lambda \in [0,1]$), a constant which exponentially discounts old observations as new ones become available, the loss function (8) at time $t$ becomes:

$$S_t(\boldsymbol{\theta}_t) = \sum_{s=1}^{t} \lambda^{t-s} \varepsilon_s^2(\boldsymbol{\theta_t}). \quad (9)$$

For details about updating formulas, properties and conditions, please see Madsen (2008, ch. 11).

## 3.2 Conditional Parametric models

In addition to the exogenous regressors $u_i$, $i = 1, \ldots, m_u$, in the model (6), the parameters can also be conditioned upon another set of external variables $\boldsymbol{v} = [v_1, \ldots v_{m_v}]$. This can be useful when the parameters are likely to be influenced by factors which do not affect $y_t$ directly. For instance, the relationship between the estimated $\alpha$-parameters for the different LIDAR cubes is likely to be affected by the wind direction due to their relative position.

Now consider the model (6) where $q = 0, d = 0$ such that

$$\boldsymbol{X}_t^T = [y_{t-1}, \ldots, y_{t-p}, u_{t-1}, \ldots, u_{t-s}] \quad (10)$$

and define a model for $Y_t$ as

$$y_t = \boldsymbol{X}_t^T \boldsymbol{\theta}(\boldsymbol{v}_t) + \varepsilon_t. \quad (11)$$

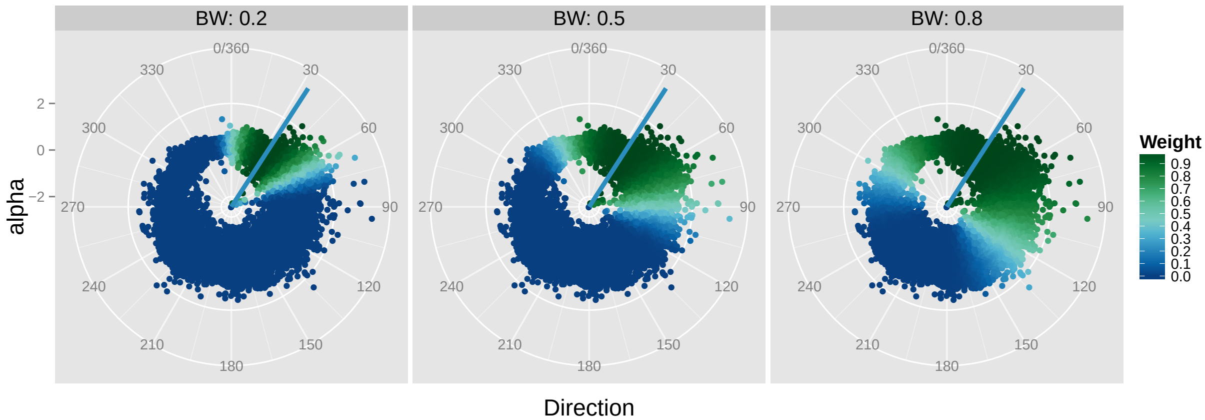



Figure 4: Example of how observation weights differ according to the relative bandwidth

This model is an ARX-model which parameters depend on another independent set of external variables $\boldsymbol{v}$. Due to what is commonly known as the curse of dimensionality, the dimension of $\boldsymbol{v}$ is usually kept low, while the dimension of $\boldsymbol{X}_t$ can be quite large.

The estimation procedure for $\boldsymbol{\theta}(\boldsymbol{v})$ is described in detail in Nielsen et al. (2000). In short, the idea behind the model structure and parameter estimation involves fitting a number of linear models (at equally many distinct fitting points of $\boldsymbol{v}$) which describe the relation between $\boldsymbol{X}_t$ and $y_t$ in the vicinity of that particular $v$.

Let $\boldsymbol{v}_j$ be a particular fitting point for $\boldsymbol{v}$ and let $p_d(\boldsymbol{v}_j)$ be a vector of terms in the corresponding polynomial of order $d$. So if $m_v = 1$ and $d = 1$, then $p_1(\boldsymbol{v}_j) = [1,\ v_{1,j}]$ (see Nielsen et al. (2000) for higher order examples). Furthermore, define

$$\boldsymbol{z}_t^T = [p_d(\boldsymbol{v}_j)X_{1,t},\ \ldots,\ p_d(\boldsymbol{v}_j)X_{p+m_u,t}]\,. \tag{12}$$

and a corresponding parameter vector $\phi_{t,j}$ as

$$\phi_t(j) = [\phi_{1,t}(j),\ \ldots,\ \phi_{k,t}(j)] \tag{13}$$

where $k$ is the number of elements in $\boldsymbol{z}_t^T$. Then the linear model

$$y_t = \boldsymbol{z}_t^T \phi_t(j) + e_t \tag{14}$$

describes $y_t$ around $\boldsymbol{v}_j$.

The parameters in $\phi_t(j)$ are found, in a recursive setting, by solving a least squares problem

$$\widehat{\phi}_t(j) = \arg\min_{phi} \sum_{s=1}^{t} (1-(1-\lambda^{t-s})w_s)\left(y_t - \boldsymbol{z}_t^T\phi_t(j)\right)^2 \tag{15}$$

where $\lambda$ is the forgetting factor as before and $w_s$ is an observation weight, assigned according to the distance between the observed $v_t$ and $\boldsymbol{v}_j$. Following Nielsen et al. (2000), the weights are assigned according to a tri-cube kernel using a relative nearest neighbor bandwidth, $\gamma$.

The effect of the choice of bandwidth on the estimation is illustrated in Figure 4. Here the $\alpha$'s for the Northern cube have been plotted against the wind direction and take colors according to the weight, $w_s$, they are assigned for three different bandwidths.

An example of the resulting parameters is shown in Figure 5. Here, the parameters of the model in (11) have been estimated, setting $p = 1$, $m_u = 1$, using the $\alpha$'s from the western and the southern cubes as $u$ and conditioning the parameters on the wind direction at $t-1$. The plots in the figure illustrate how the individual parameters vary according to the wind direction.

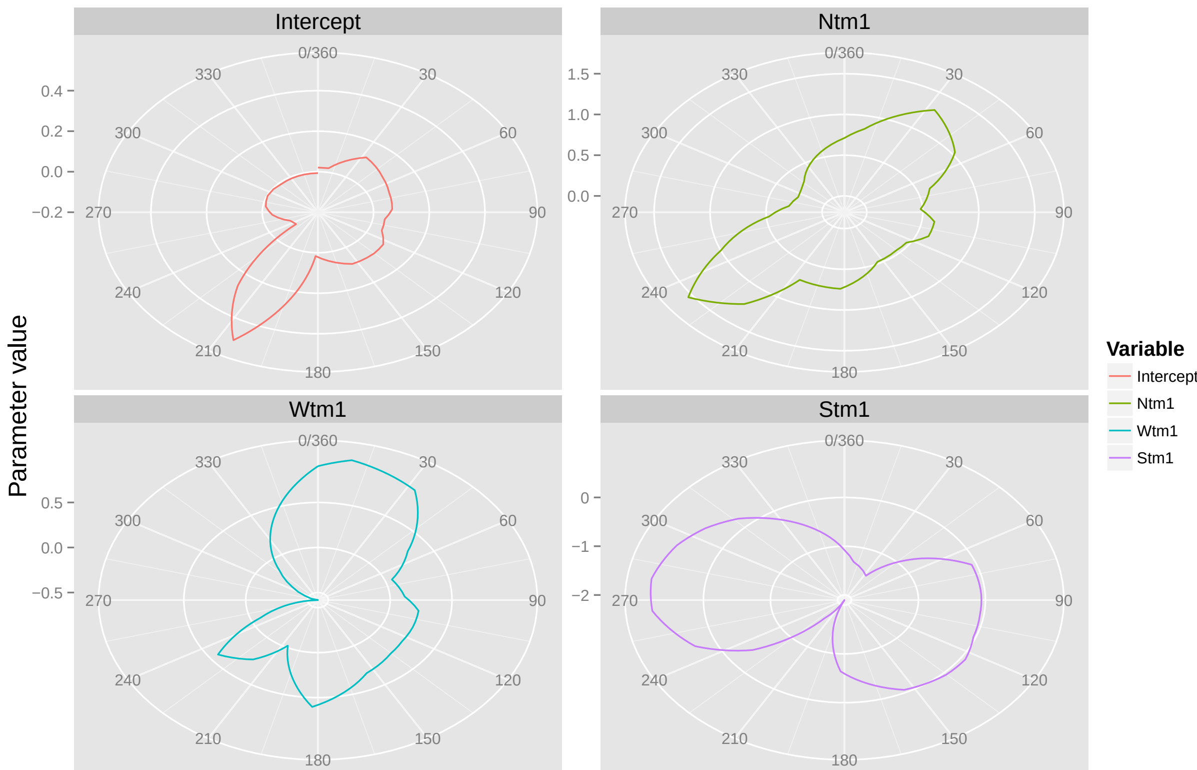



Figure 5: Example of how parameters differ according to wind direction

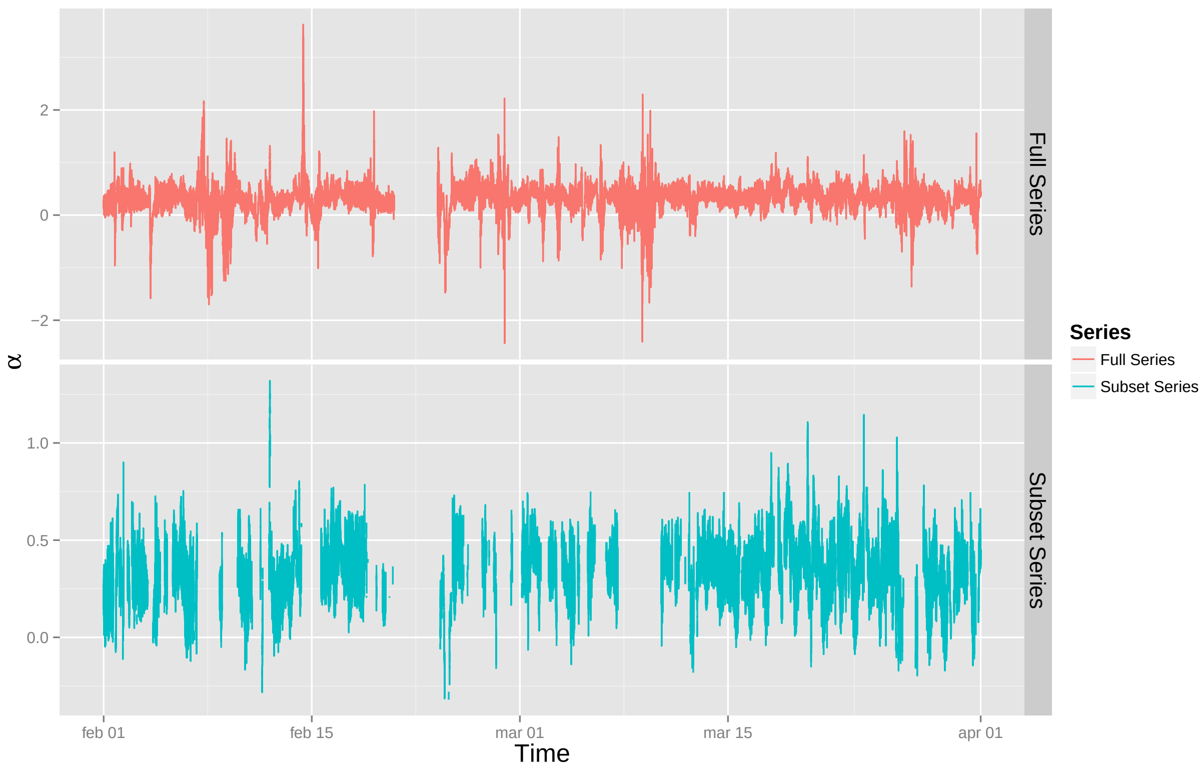



Figure 6: Time series plot of the full series and the subset series

# 4 Model Building

The different models described in the previous section, are now fitted to the data set and the prediction skill estimated in order to assess whether they are suited for online prediction of the wind profile through the $\alpha$ parameter. Since a very close correspondence between the forecasting skill for $\alpha$ and that for the wind profile itself, the models' prediction skill is reported in terms of $\alpha$. Towards the end of this section, transformation to the actual wind profile is then discussed for selected models.

The model estimation is done on the $\alpha$ for the Northern LIDAR while the other two are, in some models, used as explanatory variables. All forecasts made are single step ones. Figure 6 shows the two time series which the models are estimated on. For better illustrating the variability of the subset series, the y-axis in the bottom plot has less range.

## 4.1 Benchmarks

In order to establish a fairly simple benchmark, in addition to persistence forecasts, we start out by a simple standard ARIMA model. The model order is decided on by minimizing the Akaike Information Criteria (AIC) for the full two month example and parameters are estimated using the same data. The of course yields an unrealistically good forecasting skill of the benchmark model since information not available at the forecasting time is used for estimation. However, the estimation procedure was deemed adequate for benchmarking purposes.

Both ARIMA and persistence models are estimated for the full time series and for the subset series including only observations from periods when wind speed was above 5 $m/s$ and more than 50% of the data at 300 meters was available. The suggested model orders and the corresponding skill measures are listed in Table 1. The table shows that the ARIMA model includes a first order differentiation for both series. The model for the full series includes only a first order AR part but with a fourth order MA part while the model for the subset series only includes a fifth order AR part and no MA part. This reflects the less stability of the full series compared to the subset. Table 1 also shows that the two models perform almost equally well on a relative scale, reducing the residual standard deviation by around 10%.

## 4.2 ARIMAX by Recursive Least Squares

An alternative to differentiating the series for obtaining a series suitable for time-invariant models is to estimate the parameters recursively, with an exponential discounting of observations, as mentioned in previous section. The estimating the parameters more local in time, and thereby allowing them the vary with, the parameters adapt to the current dynamics of the system. In addition to just introducing recursively estimated parameters, model variations including lagged $\alpha$-values from the other LIDARs are also considered. Models were fitted for differentiating of order 1 and without differentiating. It turns out that allowing the model parameters to vary over time makes differentiating not necessary and thus the model has the form:

$$\alpha_t^{(N)} = \phi_{0,t} + \sum_{k=1}^{p} \phi_{k,t}\alpha_{t-k}^{(N)} + \sum_{k=1}^{q} \theta_{k,t}\varepsilon_{t-k} + \sum_{k=1}^{m_1} \omega_{k,t}^{(W)}\alpha_{t-k}^{(W)} + \sum_{k=1}^{m_2} \omega_{k,t}^{(S)}\alpha_{t-k}^{(S)} + \varepsilon_t \tag{16}$$

where $\alpha_t^{(N)}$, $\alpha_t^{(W)}$ and $\alpha_t^{(S)}$ are the $\alpha$'s, estimated at time $t$ for the northern, western and southern LIDAR cubes respectively.

The model order and forgetting factor tried were every possible combination of:

$$p \in \{1,\ \dots,\ 5\}, \quad q \in \{0,\ \dots,\ 4\}, \quad m_1, m_2 \in \{0,\ 1,\ 2\}$$
$$\lambda \in \{0.8,\ 0.889,\ 0.9384,\ 0.9658,\ 0.981,\ 0.9895,\ 0.9942,\ 0.9968,\ 0.9982,\ 0.999\}$$

Table 2 lists the model order and forgetting factor that yielded the best forecasting skill along with the corresponding performance measures. Disappointingly, only a marginally improved forecasting skill is obtained, compared to the time invariant models. It is interesting to see though that the MA-term becomes obsolete when parameter estimation is done in a recursive fashion though.

## 4.3 Conditional Parametric ARX model

Now the conditional parametric model described in the previous section is used for forecasting $\alpha_t^{(N)}$. Due to quite excessive computation time for each model setup and the number of tuning parameters, only the two

Table 1: Model orders of the static ARIMA$(p, d, q)$ models.

| | | Persistence | | ARIMA | | | |
|---|---|---|---|---|---|---|---|
| | $(p, d, q)$ | RMSE | MAE | RMSE | MAE | RMSSE | MASE |
| Full Series | (1, 1, 4) | 0.1093 | 0.0708 | 0.0988 | 0.0639 | 0.9043 | 0.9026 |
| Subset Series | (5, 1, 0) | 0.0715 | 0.0533 | 0.0640 | 0.0477 | 0.8943 | 0.8947 |

Table 2: The forecasting skill of the best performing RLS-ARMAX models.

| | $\lambda$ | $(p, q, m_1, m_2)$ | RMSE | MAE | RMSSE | MASE |
|---|---|---|---|---|---|---|
| Full Series | 0.9895 | (3,0,0,0) | 0.0969 | 0.0625 | 0.8866 | 0.8827 |
| Subset Series | 0.9895 | (1,1,0,0) | 0.0630 | 0.0471 | 0.8805 | 0.8833 |

model orders were tried for 9 different combinations of the meta parameters, $\lambda$ and $\gamma$:

$$\text{Order 1: } (p, m_1, m_2) = (1, 1, 1), \quad \text{Order 2: } (p, m_1, m_2) = (2, 0, 0),$$
$$\lambda \in \{0.8,\ 0.9,\ 0.99\}, \quad \gamma \in \{0.2,\ 0.5,\ 0.8\}$$

Both differentiation of order 0 and 1 were tried. It turns out that for both model orders tried and for both differentiation orders, the best performance was observed for $(\gamma,\ \lambda) = (0.8,\ 0.8)$. The corresponding performance measures are reported in Tables 3 - 4.

Significantly improved forecasting skill is observed in both cases, compared to the previous models. However, a forgetting factor of 0.8, which corresponds to 5 effective observations in the estimation, is disturbingly low. And even though the effective memory is somewhat higher due to the combination with the wind directional weight, there is still a risk of unstable parameter estimates and model performance with such a low $\lambda$, even though not observed in this short period. This low value of $\lambda$ indicates that some influential factors are not taken into account in the model. However, due to that only a limited number of variables were available for the study, the inclusion of other variables was not consider.

## 4.4 Three Step Model

In an attempt to obtain a model with forecasting skill similar to what was seen in for the conditional parametric models, a model describing $\alpha_t^{(N)}$ in three independent steps is constructed. The idea is to use exponential smoothing and RLS-ARIMA model to filter out the slow drift in the data on one hand and the fast variations on the other hand. Then a conditional parametric model is added as a third and final step. Each model describes the residuals from the previous step.

The first step is an exponential smoothing which tracks the slow drift in $\alpha_t^{(N)}$, caused by e.g. seasonal variations in weather and surface roughness. The model can be written as:

$$\mu_t^{\alpha_t^{(N)}} = \mu_{t-1}^{\alpha_t^{(N)}} + \left(1 - \lambda^{(ES)}\right)\left(\alpha_t^{(N)} - \mu_{t-1}^{\alpha_t^{(N)}}\right). \tag{17}$$

and a single-step prediction is found as simply as:

$$\widehat{\alpha}_{t|t-1}^{(N)} = \mu_{t-1}^{\alpha_t^{(N)}} \tag{18}$$

For the RLS-ARIMA model, three different model orders are tried. The first one is found as the one that minimizes the AIC for a time-invariant model for the residual series from the exponential smoothing model.

Table 3: Forecasting skill of the conditional parametric model without differentiation

| | $(\gamma, \lambda)$ | Order | RMSE | MAE | RMSSE | MASE |
|---|---|---|---|---|---|---|
| Full Series | (0.8, 0.8) | 1 | 0.0323 | 0.0182 | 0.4563 | 0.1661 |
| | (0.8, 0.8) | 2 | 0.0414 | 0.0244 | 0.5840 | 0.2233 |
| Subset Series | (0.8, 0.8) | 1 | 0.0210 | 0.0137 | 0.2935 | 0.2565 |
| | (0.8, 0.8) | 2 | 0.0270 | 0.0184 | 0.3774 | 0.3449 |

Table 4: Forecasting skill of the conditional parametric model with first order differentiation

| | $(\gamma, \lambda)$ | Order | RMSE | MAE | RMSSE | MASE |
|---|---|---|---|---|---|---|
| Full Series | (0.8, 0.8) | 1 | 0.0245 | 0.0133 | 0.2167 | 0.0771 |
| | (0.8, 0.8) | 2 | 0.0414 | 0.0244 | 0.3655 | 0.1413 |
| Subset Series | (0.8, 0.8) | 1 | 0.0160 | 0.0101 | 0.1390 | 0.1170 |
| | (0.8, 0.8) | 2 | 0.0270 | 0.0184 | 0.2340 | 0.2138 |

This yields $(p, d, q) = (5, 0, 1)$ for the full series and $(p, d, q) = (4, 0, 1)$ for the subset series. In addition a models of order $(2, 0, 1)$ and $(2, 0, 0)$ are also tried.

Tables 5 - 6 list the forecasting skill of the three step model, step by step, on an absolute and relative scale respectively. Disappointingly, $\lambda = 0.8$ still yields the best performance and the three step model is not capable of outperforming the previously discussed conditional parametric model. Thus it must be concluded that a more elaborate model, accounting for more external factors is required in order to increase the optimal forgetting factor of the conditional parametric model.

Table 5: Absolute forecasting skill measures for the each step of the three step model.

| | | | | | ES | | RLS | | CP | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\lambda^{(ES)}$ | $\lambda^{(RLS)}$ | $(p, d, q)$ | $(\lambda^{(CP)}, \gamma)$ | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| | — | 0.99 | (2,0,1) | (0.8,0.8) | — | — | 0.0625 | 0.0972 | 0.0448 | 0.0261 |
| Full Series | 0.9999 | 0.99 | (2,0,1) | (0.8,0.8) | 0.2242 | 0.1556 | 0.0625 | 0.0972 | 0.0448 | 0.0261 |
| | 0.9900 | 0.99 | (2,0,1) | (0.8,0.8) | 0.1504 | 0.0984 | 0.0625 | 0.0972 | 0.0448 | 0.0260 |
| | — | 0.99 | (2,0,1) | (0.8,0.8) | — | — | 0.0466 | 0.0622 | 0.0286 | 0.0193 |
| Subset Series | 0.9999 | 0.99 | (2,0,1) | (0.8,0.8) | 0.1378 | 0.1055 | 0.0466 | 0.0622 | 0.0286 | 0.0193 |
| | 0.9900 | 0.99 | (2,0,1) | (0.8,0.8) | 0.0929 | 0.0698 | 0.0465 | 0.0621 | 0.0286 | 0.0193 |

Table 6: Relative forecasting skill measures for the each step of the three step model.

| | | | | | ES | | RLS | | CP | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\lambda^{(ES)}$ | $\lambda^{(RLS)}$ | $(p, d, q)$ | $(\lambda^{(CP)}, \gamma)$ | RMSSE | MASE | RMSSE | MASE | RMSSE | MASE |
| | — | 0.99 | (2,0,1) | (0.8,0.8) | — | — | 0.5723 | 1.3733 | 0.4097 | 0.3682 |
| Full Series | 0.9999 | 0.99 | (2,0,1) | (0.8,0.8) | 2.0517 | 2.1975 | 0.5723 | 1.3733 | 0.4097 | 0.3682 |
| | 0.9900 | 0.99 | (2,0,1) | (0.8,0.8) | 1.3763 | 1.3902 | 0.5719 | 1.3722 | 0.4095 | 0.3678 |
| | — | 0.99 | (2,0,1) | (0.8,0.8) | — | — | 0.6509 | 1.1677 | 0.4003 | 0.3626 |
| Subset Series | 0.9999 | 0.99 | (2,0,1) | (0.8,0.8) | 1.9262 | 1.9799 | 0.6509 | 1.1678 | 0.4003 | 0.3626 |
| | 0.9900 | 0.99 | (2,0,1) | (0.8,0.8) | 1.2982 | 1.3090 | 0.6502 | 1.1658 | 0.3994 | 0.3618 |

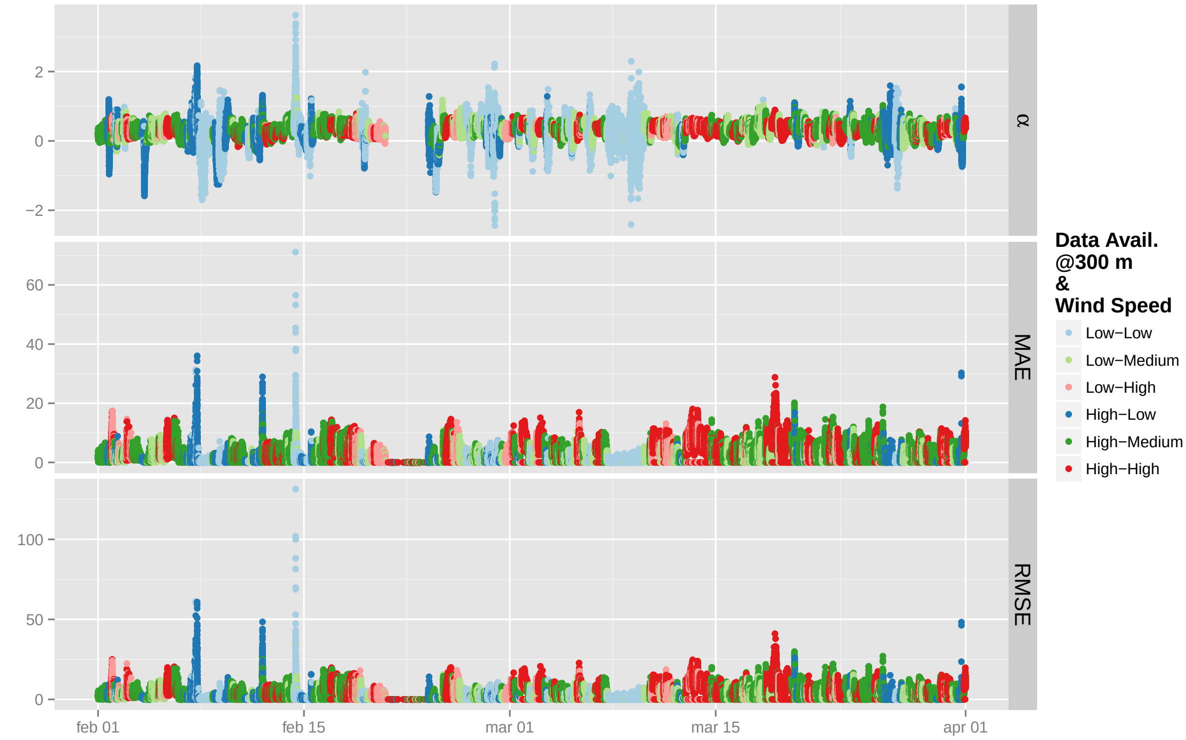



Figure 7: Time series plot of the 30 second $RMSE$ and $MAE$ for the $\alpha$-estimates.

# 5 Conclusions

Several statistical models have been tried for short-term prediction of wind profiles. The trials have been successful in the sense that models, which are able to provide fairly accurate forecasts, have been identified as conditional parametric ARX models. However, there are several issues that need to be addressed before the experiment can be regarded as a successful one.

First of all, the actual estimation of the $\alpha$-parameter deserves a closer look. The abnormal values found, mainly during periods of low wind speed, indicate that a more elaborate method for describing the wind profile should be considered. As shown in Figure 7, the accuracy of the $\alpha$-estimates are poorest during the low wind periods. Regardless of how the wind profiles are described, the distinctive behavior of $\alpha$ for the low wind observations indicate some extra measures have to be taken to model these periods properly. In this regard, regime switching models of some kind come to mind as an appealing approach.

For the $\alpha$ estimates obtained during this study, there are still some possible extensions that might be worth pursuing. For instance, one can imagine that filtering out the short-term variations by the exponential smoothing step of the three step model (instead of the long-term drift) might helpful for increasing the optimal memory of the conditional parametric model. Likewise, adding a global (or local) MA-part to the conditional parametric model might have the same effect.

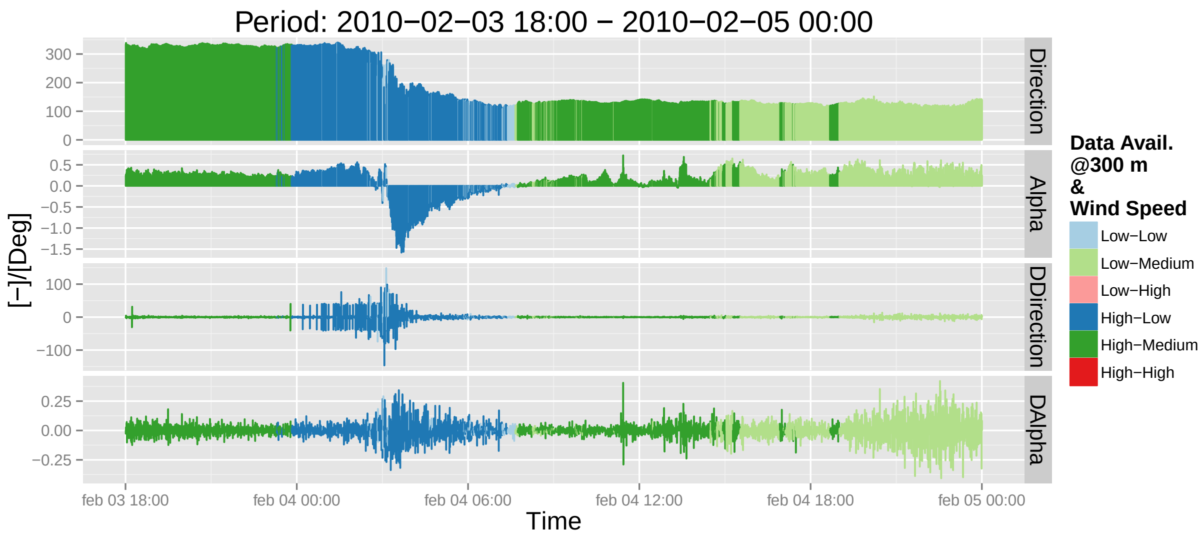
# References

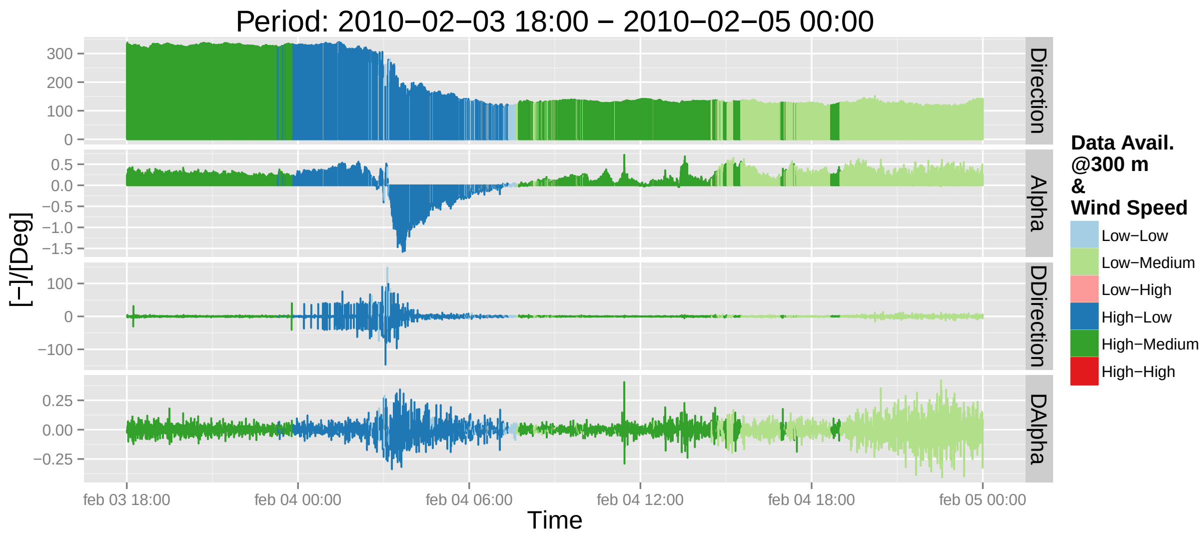
Henrik Madsen. Time Series Analysis. Chapman & Hall / CRC, London, UK, 2008.

Jakob Mann, Mike Courtney, Poul Hummelshøj, and Peter Hjuler Jensen. Undersøgelse af vindforhold ved det kommende testcenter ved østerild. Technical report, Risø DTU, 2010.

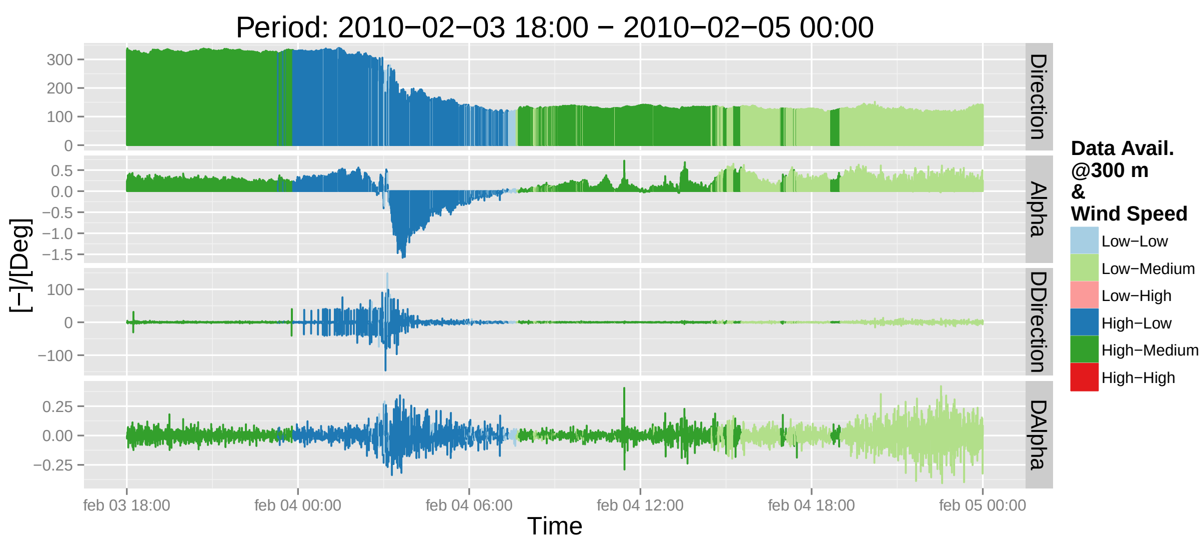
Henrik Aa. Nielsen, Torben S. Nielsen, Alfred K. Joensen, Henrik Madsen, and Jan Holst. Tracking time-varying coefficient functions. International Journal of Adaptive Control and Signal Processing, 14(8): 813–828, 2000.

# A Further examples of unstable $\alpha$-parameters

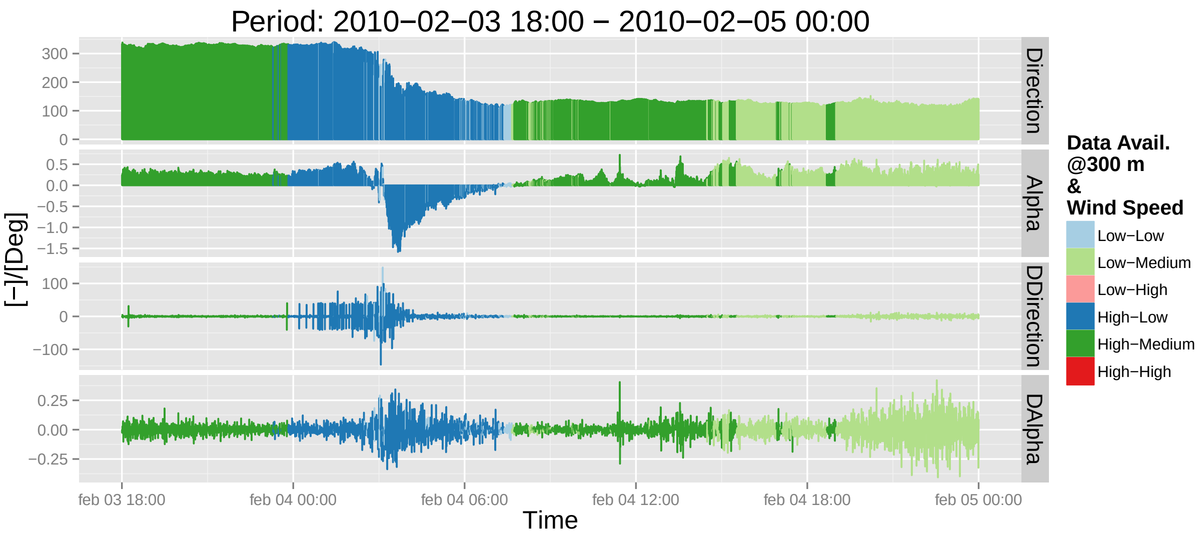


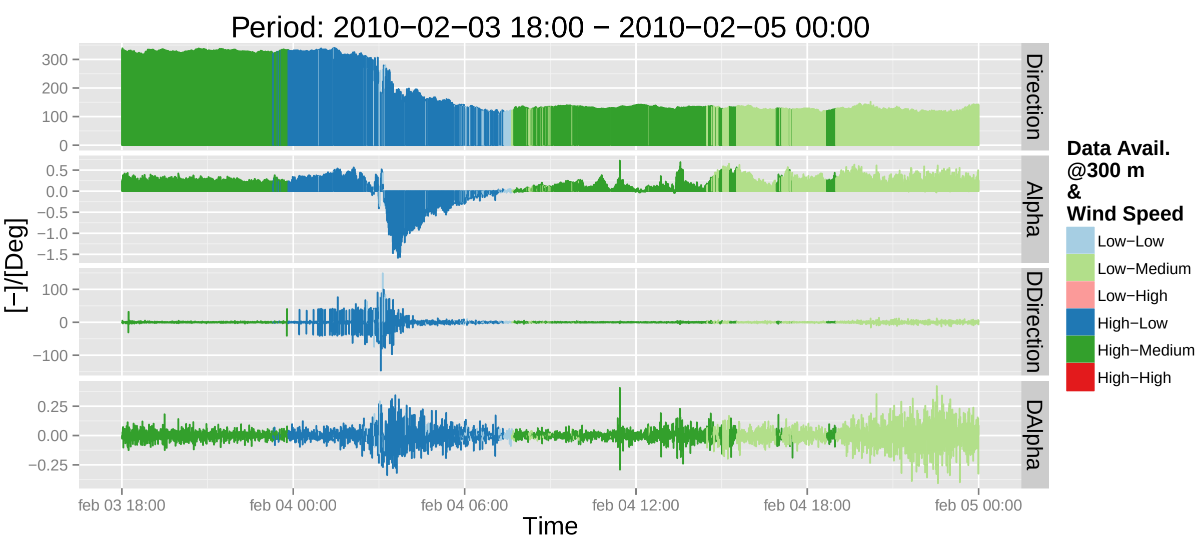


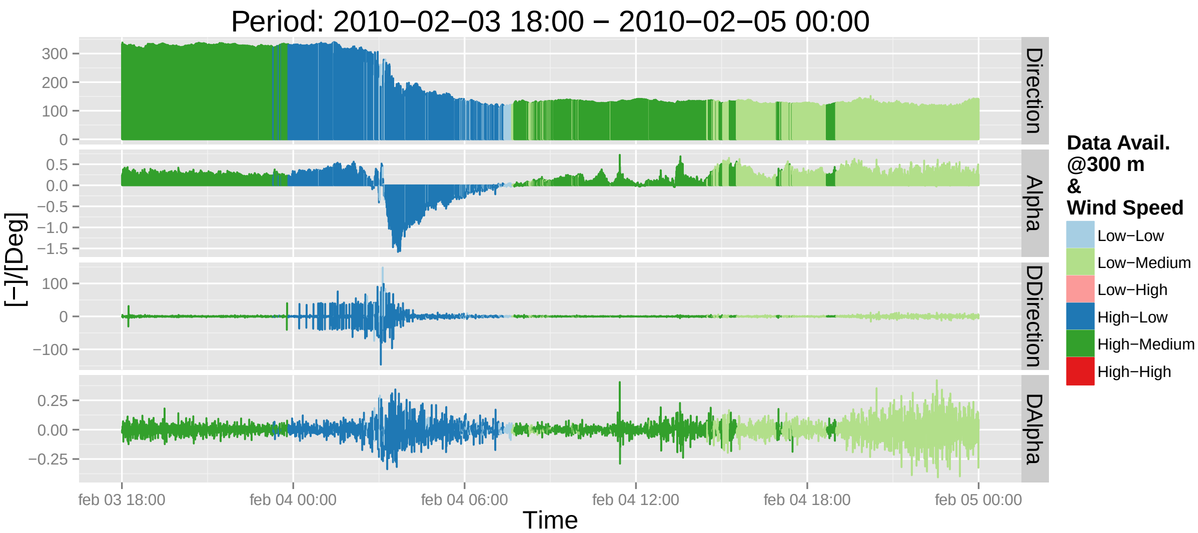
Period: 2010−02−03 18:00 − 2010−02−05 00:00
Direction
Alpha
DDirection
DAlpha
[−]/[Deg]
Time
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Data Avail. @300 m & Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High

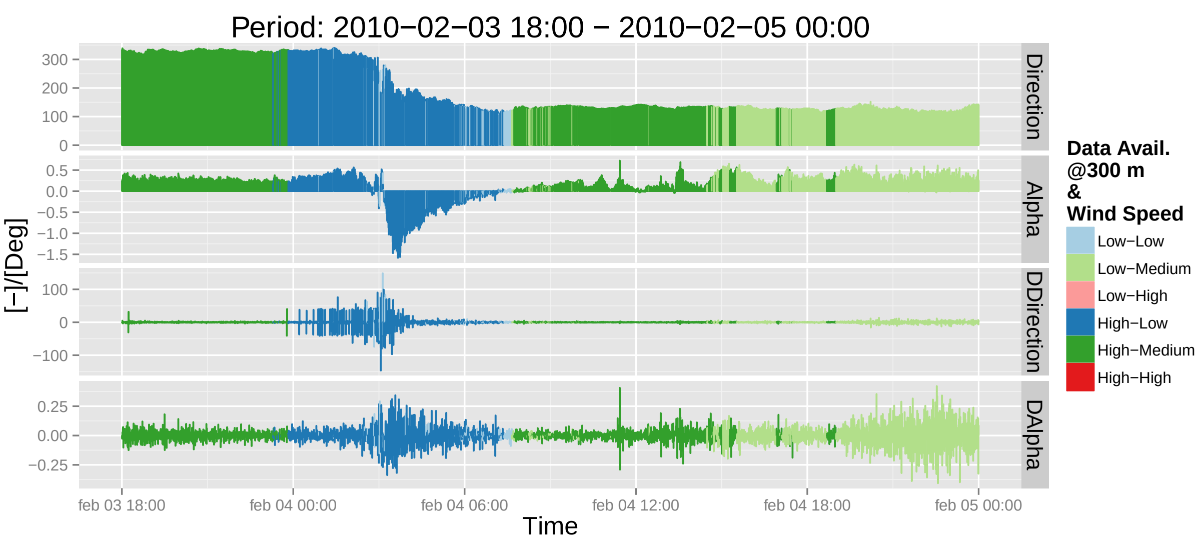



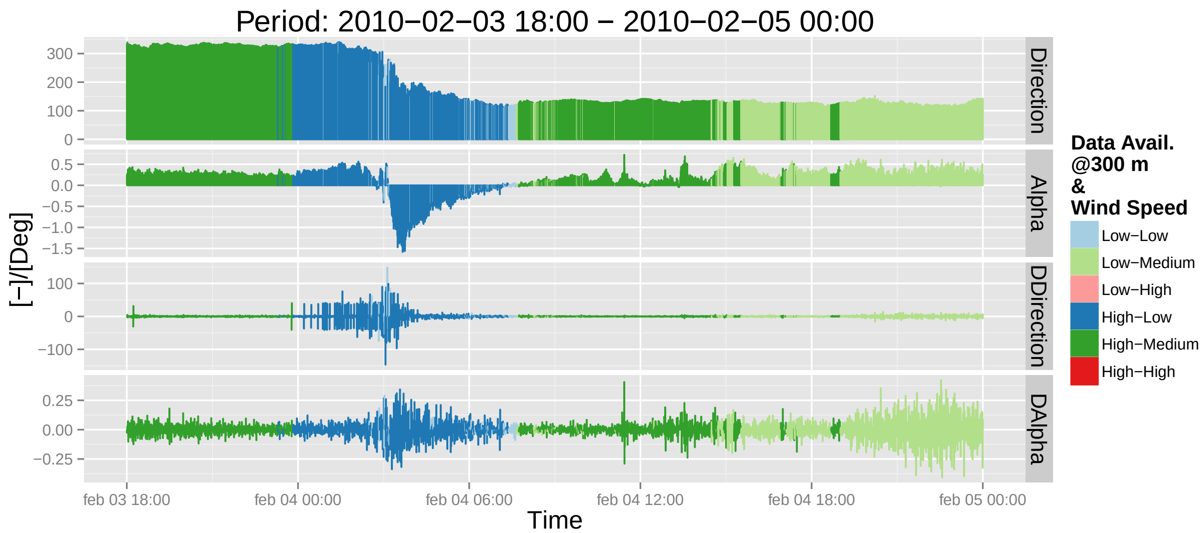
Period: 2010−02−03 18:00 − 2010−02−05 00:00
Direction
Alpha
DDirection
DAlpha
[−]/[Deg]
Time
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Data Avail. @300 m & Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High

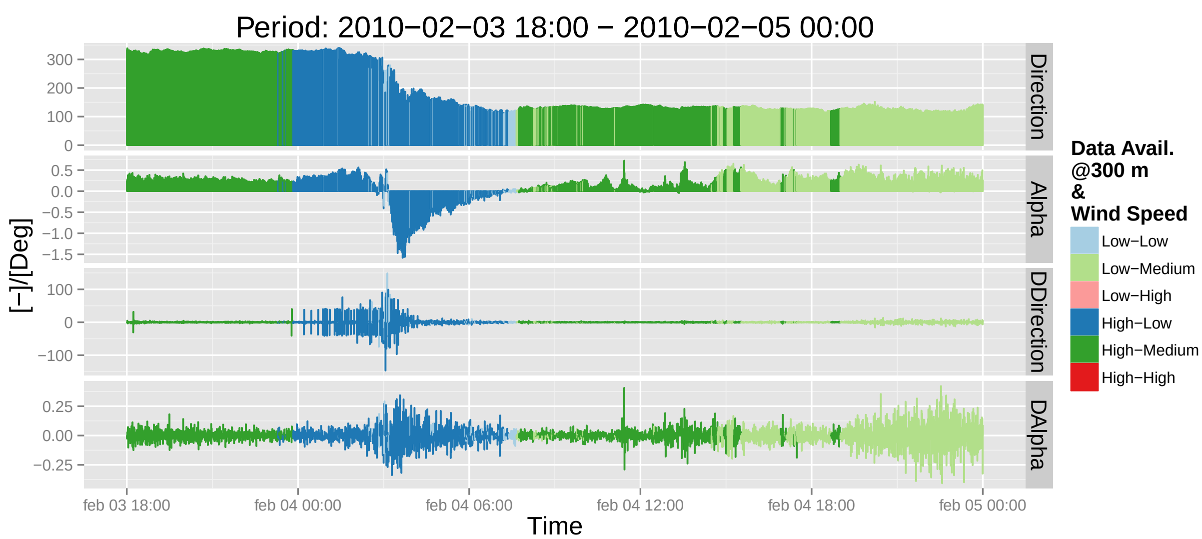

Period: 2010−02−03 18:00 − 2010−02−05 00:00
300
200
100
0
0.5
0.0
−0.5
−1.0
−1.5
100
0
−100
0.25
0.00
−0.25
[−]/[Deg]
Direction
Alpha
DDirection
DAlpha
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Time
Data Avail.
@300 m
&
Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High

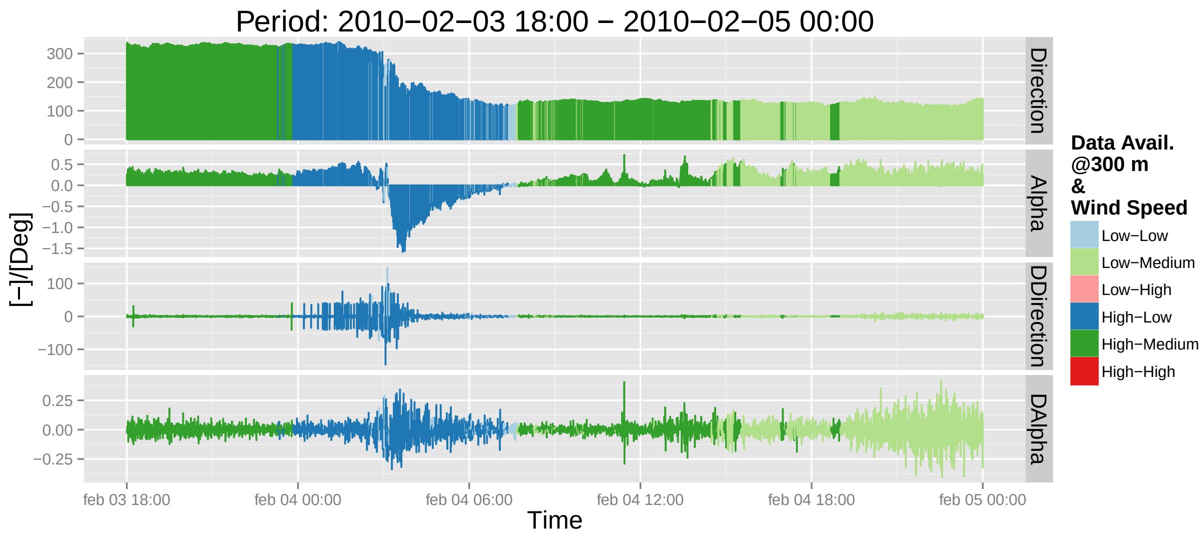

Period: 2010−02−03 18:00 − 2010−02−05 00:00
300
200
100
0
0.5
0.0
−0.5
−1.0
−1.5
100
0
−100
0.25
0.00
−0.25
[−]/[Deg]
Direction
Alpha
DDirection
DAlpha
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Time
Data Avail.
@300 m
&
Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High


Period: 2010−02−03 18:00 − 2010−02−05 00:00
Direction
Alpha
DDirection
DAlpha
[−]/[Deg]
300
200
100
0
0.5
0.0
−0.5
−1.0
−1.5
100
0
−100
0.25
0.00
−0.25
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Time
Data Avail.
@300 m
&
Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High

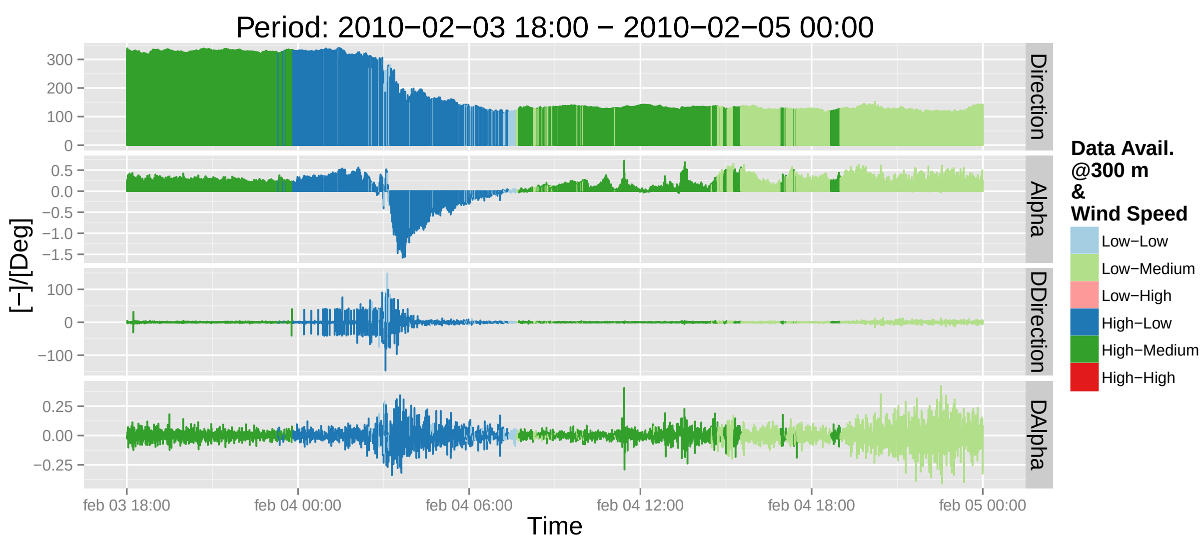

Period: 2010−02−03 18:00 − 2010−02−05 00:00
Direction
Alpha
DDirection
DAlpha
[−]/[Deg]
300
200
100
0
0.5
0.0
−0.5
−1.0
−1.5
100
0
−100
0.25
0.00
−0.25
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Time
Data Avail.
@300 m
&
Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High


Period: 2010−02−03 18:00 − 2010−02−05 00:00
Direction
Alpha
DDirection
DAlpha
[−]/[Deg]
Time
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Data Avail.
@300 m
&
Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High

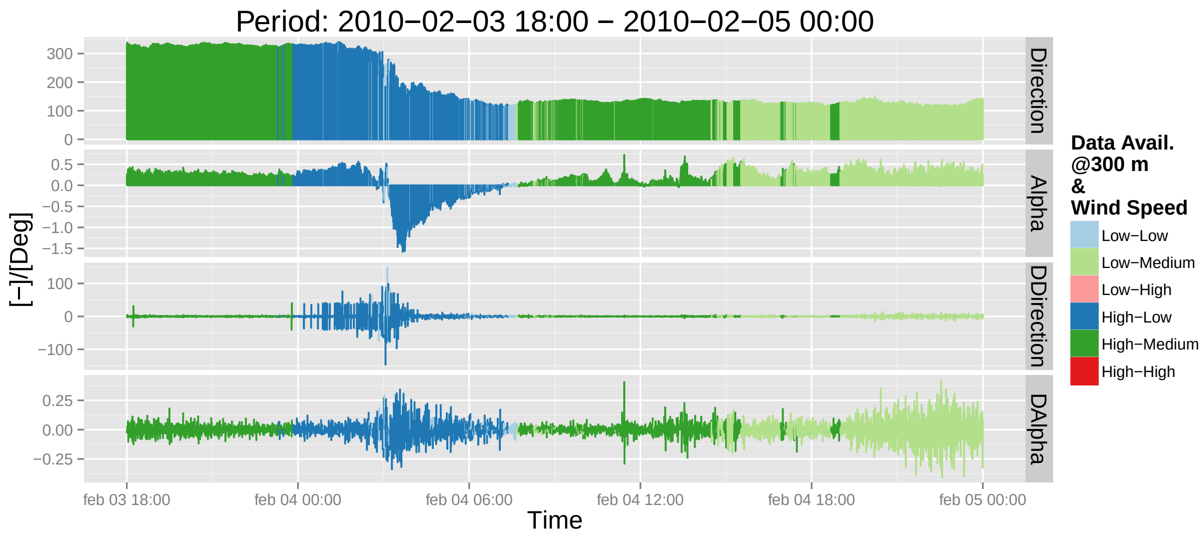

Period: 2010−02−03 18:00 − 2010−02−05 00:00
Direction
Alpha
DDirection
DAlpha
[−]/[Deg]
Time
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Data Avail.
@300 m
&
Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High

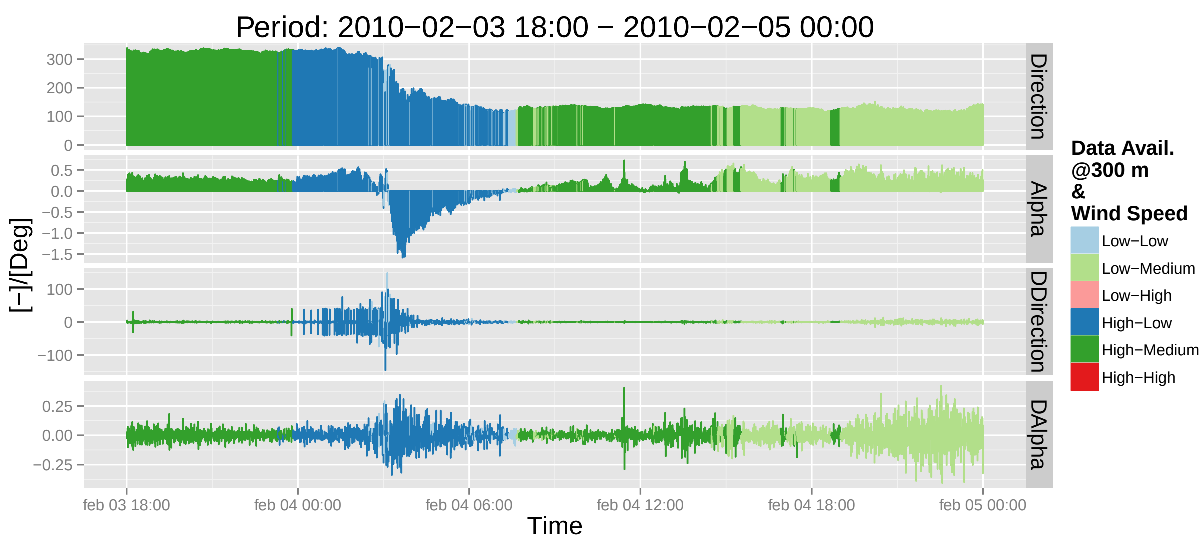


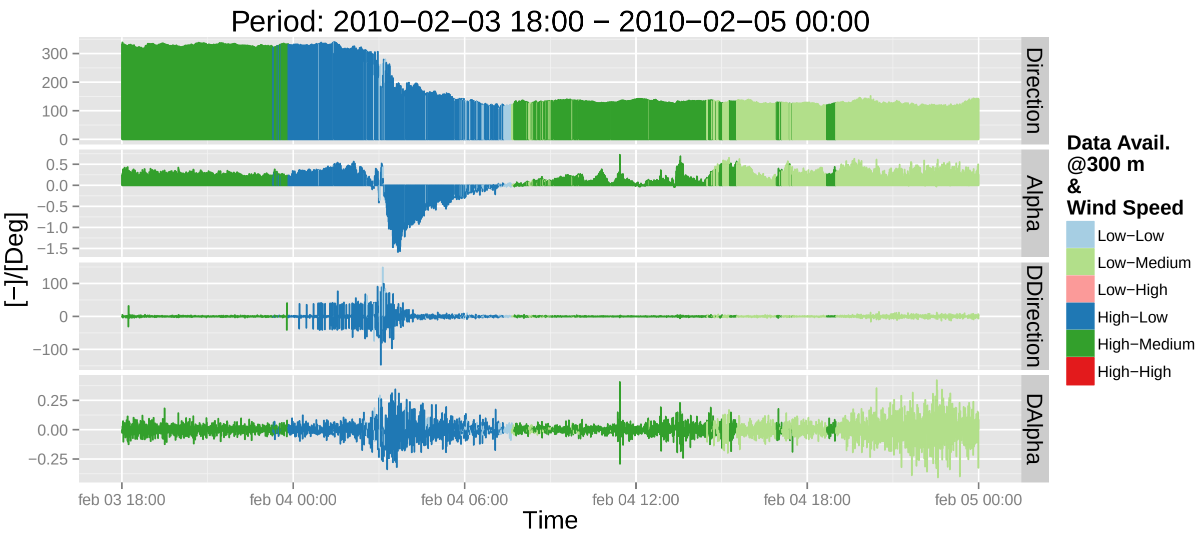
Period: 2010−02−03 18:00 − 2010−02−05 00:00
[−]/[Deg]
Direction
Alpha
DDirection
DAlpha
Time
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Data Avail.
@300 m
&
Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High

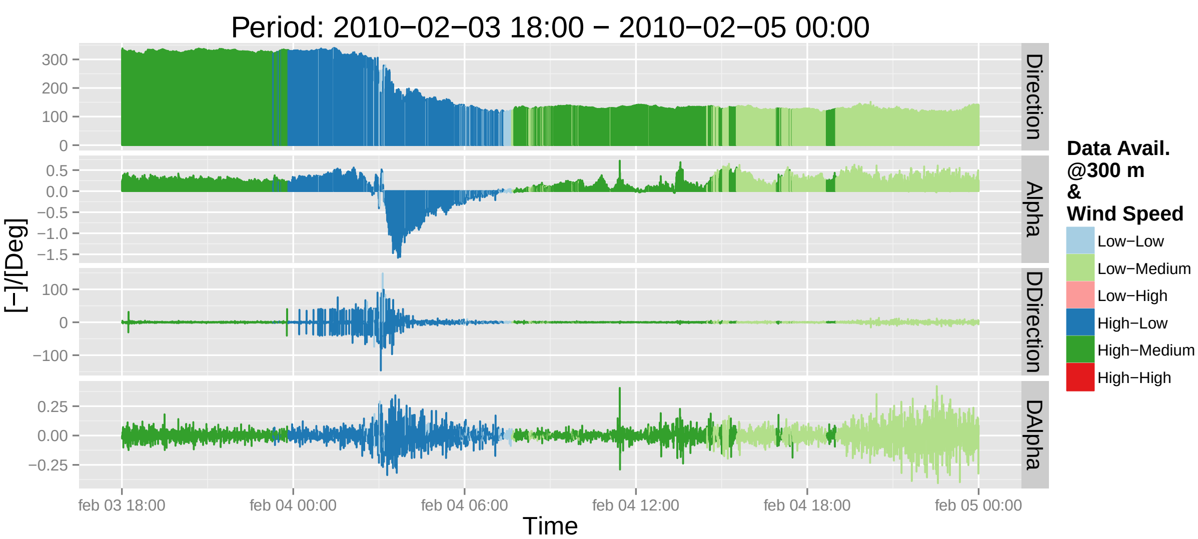
Period: 2010−02−03 18:00 − 2010−02−05 00:00
[−]/[Deg]
Direction
Alpha
DDirection
DAlpha
Time
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Data Avail.
@300 m
&
Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High


Period: 2010−02−03 18:00 − 2010−02−05 00:00
[−]/[Deg]
Direction
Alpha
DDirection
DAlpha
300
200
100
0
0.5
0.0
−0.5
−1.0
−1.5
100
0
−100
0.25
0.00
−0.25
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Time
Data Avail. @300 m & Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High


Period: 2010−02−03 18:00 − 2010−02−05 00:00
[−]/[Deg]
Direction
Alpha
DDirection
DAlpha
300
200
100
0
0.5
0.0
−0.5
−1.0
−1.5
100
0
−100
0.25
0.00
−0.25
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00
Time
Data Avail. @300 m & Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High



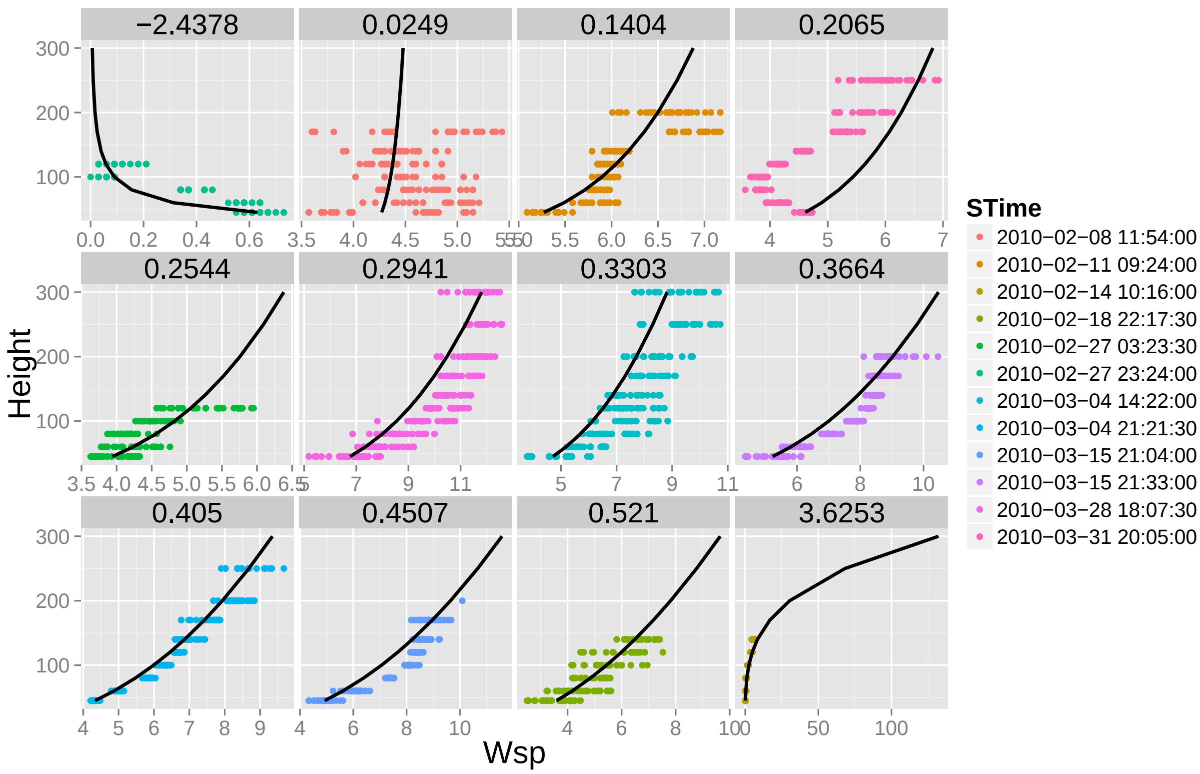
Period: 2010−02−03 18:00 − 2010−02−05 00:00
Direction
Alpha
DDirection
DAlpha
[−]/[Deg]
Time
Data Avail.
@300 m
&
Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00

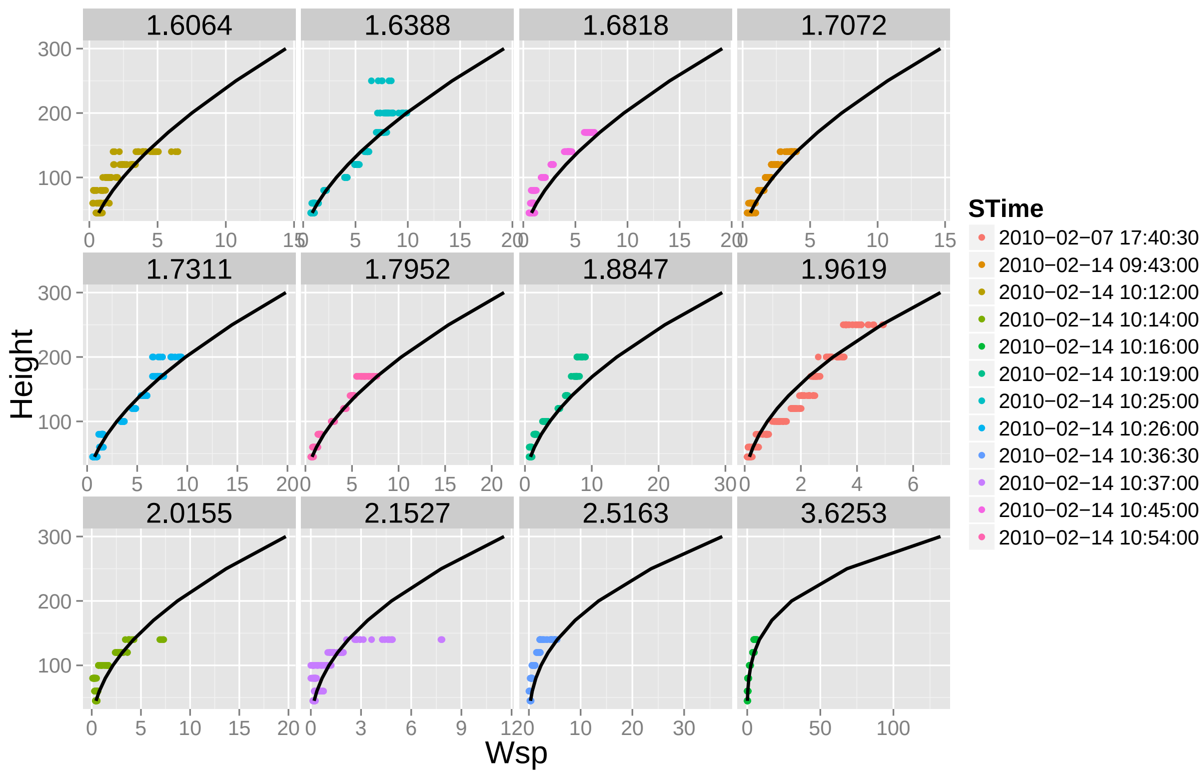



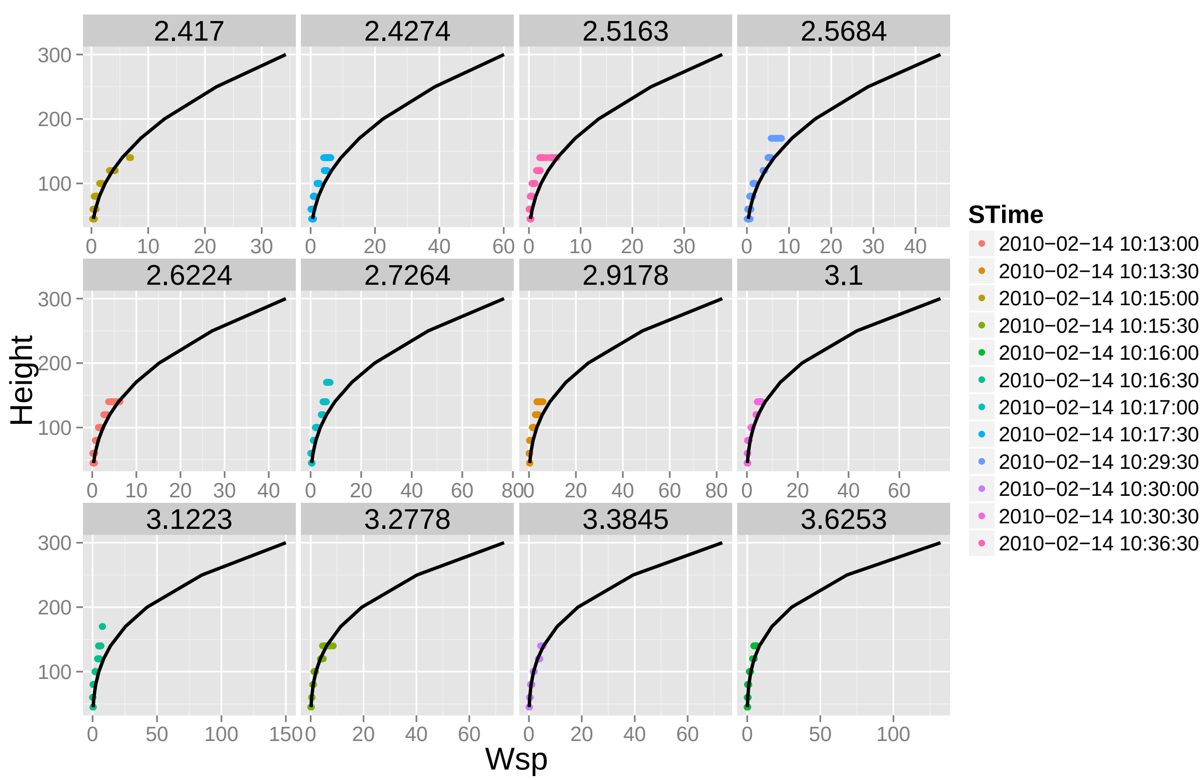
Period: 2010−02−03 18:00 − 2010−02−05 00:00
Direction
Alpha
DDirection
DAlpha
[−]/[Deg]
Time
Data Avail.
@300 m
&
Wind Speed
Low−Low
Low−Medium
Low−High
High−Low
High−Medium
High−High
feb 03 18:00
feb 04 00:00
feb 04 06:00
feb 04 12:00
feb 04 18:00
feb 05 00:00

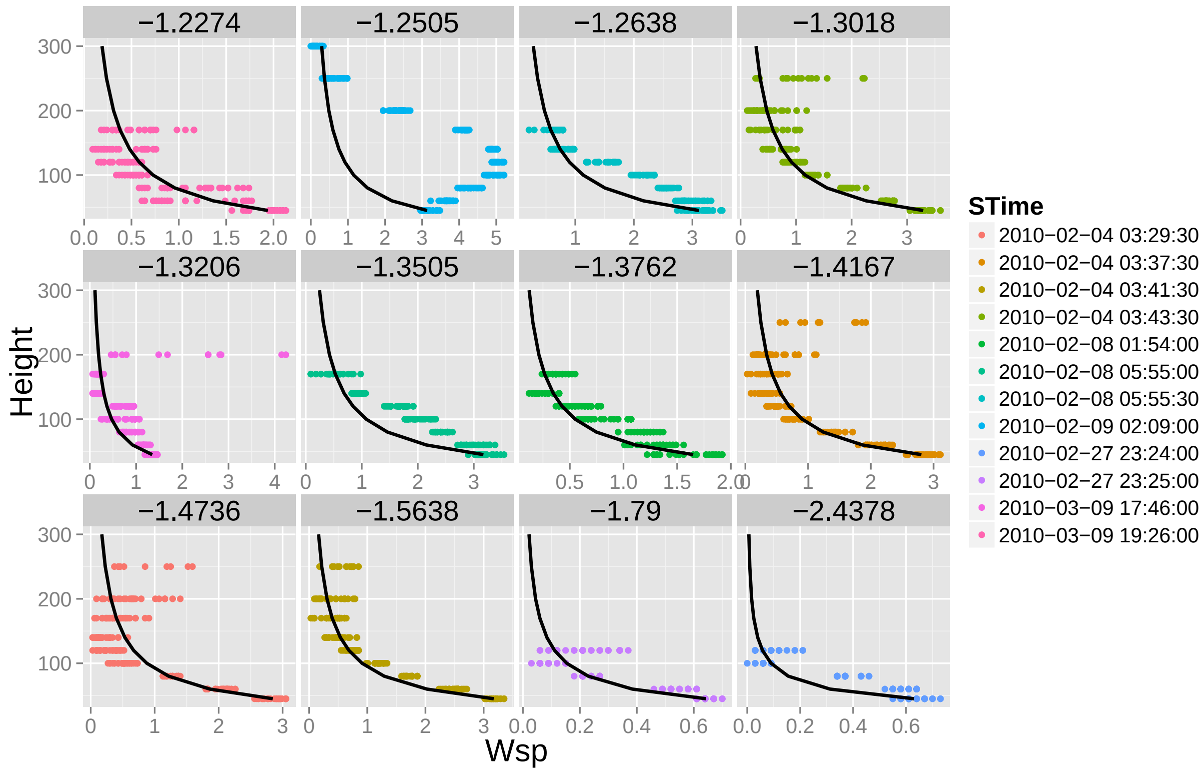
# B Examples of varying fitting quality

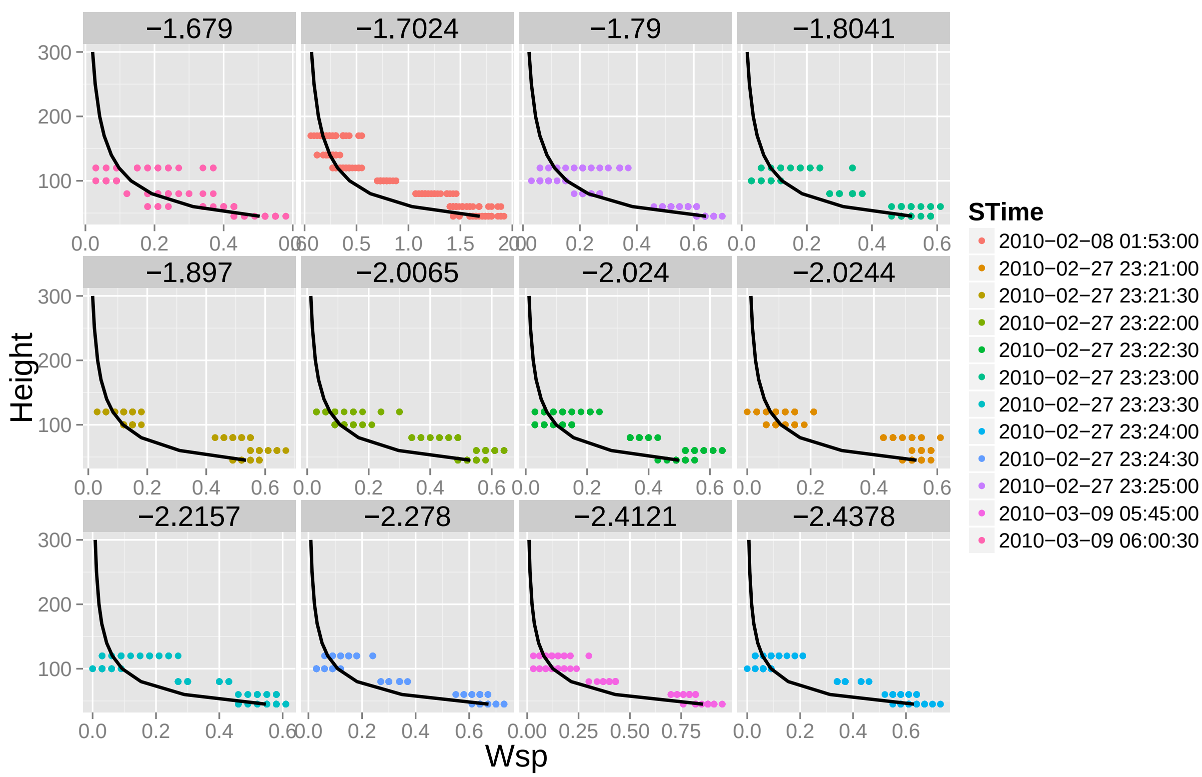
The following plots show the estimated wind profile and the corresponding data for a range of different $\alpha$-values. The estimated $\alpha$ is given in the header of each plot while the time of the observation is given in the legend. As can be seen, the fit seems rather decent for $\alpha$'s within the normal range while low wind speed and missing data at high altitutes are frequent for the abnormal values.




1.6064
1.6388
1.6818
1.7072
1.7311
1.7952
1.8847
1.9619
2.0155
2.1527
2.5163
3.6253
Height
Wsp
STime
2010−02−07 17:40:30
2010−02−14 09:43:00
2010−02−14 10:12:00
2010−02−14 10:14:00
2010−02−14 10:16:00
2010−02−14 10:19:00
2010−02−14 10:25:00
2010−02−14 10:26:00
2010−02−14 10:36:30
2010−02−14 10:37:00
2010−02−14 10:45:00
2010−02−14 10:54:00


2.417
2.4274
2.5163
2.5684
2.6224
2.7264
2.9178
3.1
3.1223
3.2778
3.3845
3.6253
Height
Wsp
STime
2010−02−14 10:13:00
2010−02−14 10:13:30
2010−02−14 10:15:00
2010−02−14 10:15:30
2010−02−14 10:16:00
2010−02−14 10:16:30
2010−02−14 10:17:00
2010−02−14 10:17:30
2010−02−14 10:29:30
2010−02−14 10:30:00
2010−02−14 10:30:30
2010−02−14 10:36:30


−1.2274
−1.2505
−1.2638
−1.3018
−1.3206
−1.3505
−1.3762
−1.4167
−1.4736
−1.5638
−1.79
−2.4378
Height
Wsp
STime
2010−02−04 03:29:30
2010−02−04 03:37:30
2010−02−04 03:41:30
2010−02−04 03:43:30
2010−02−08 01:54:00
2010−02−08 05:55:00
2010−02−08 05:55:30
2010−02−09 02:09:00
2010−02−27 23:24:00
2010−02−27 23:25:00
2010−03−09 17:46:00
2010−03−09 19:26:00


−1.679
−1.7024
−1.79
−1.8041
−1.897
−2.0065
−2.024
−2.0244
−2.2157
−2.278
−2.4121
−2.4378
Height
Wsp
STime
2010−02−08 01:53:00
2010−02−27 23:21:00
2010−02−27 23:21:30
2010−02−27 23:22:00
2010−02−27 23:22:30
2010−02−27 23:23:00
2010−02−27 23:23:30
2010−02−27 23:24:00
2010−02−27 23:24:30
2010−02−27 23:25:00
2010−03−09 05:45:00
2010−03−09 06:00:30